OE FORM 6000, 2/69

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

ERIC ACC. NO.				
ED 039 392				
CH ACC. NO.	P.A.	PUBL. DATE	ISSUE	
AA 000 570		67	RIEOCT70	

IS DOCUMENT COPYRIGHTED? YES ☐ NO ☒
ERIC REPRODUCTION RELEASE? YES ☐ NO ☒
LEVEL OF AVAILABILITY I ☒ II ☐ III ☐

AUTHOR
Goslin, David A.

TITLE
Criticism of Standardized Tests and Testing.

SOURCE CODE	INSTITUTION (SOURCE)
QPX15900	College Entrance Examination Board, New York, N.Y.
SP. AG. CODE	SPONSORING AGENCY

EDRS PRICE	CONTRACT NO.	GRANT NO.
0.25;1.80		
REPORT NO.		BUREAU NO.

AVAILABILITY

JOURNAL CITATION

DESCRIPTIVE NOTE
34p.

DESCRIPTORS *Standardized Tests; *Educational Testing; *Test Validity; *Test Reliability; *Evaluation Criteria; Predictive Ability (Testing); Policy Formation; Social Influences; Test Results; Personality Tests; Testing Problems; Aptitude Tests; Predictive Validity; Achievement Tests; Testing

IDENTIFIERS

ABSTRACT This review summarizes the major criticisms of standardized tests and testing and attempts an evaluation of the validity of the criticisms. Viewed in the context of type of test, manner of its use, and assumptions regarding its validity, the paper discusses: (1) the imperfect-prediction problem, where predictions of performance are invalidated by small differences between members of a group being tested, by assumption of existence of linear relationship between qualities tested and later occupational success, and by difficulties raised by presence of qualities not measured by standardized tests; (2) the rigid reliance on standardized test scores; (3) inherited versus acquired abilities; and, (4) the contribution of tests to their own validity of functioning as self-fulfilling prophecies. Criticisms that are independent of test validity are also examined: the difficulty for children to acquire the ability to deal with issues on which there is no clear right or wrong answer; the contention that teachers and schools are being evaluated along with the children; the general impact of tests on curriculums; and, the implications for the groups when tests are used to sort individuals into groups. Concluding remarks focus on the consideration of the social effects of tests, and on the need to develop rational and systematic policies regarding use of tests with the culturally disadvantaged, dissemination of test results, and the problem of invasion of privacy. (RJ)

ED 039 392

Criticism of Standardized Tests and Testing

David A. Goslin
Russell Sage Foundation

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

AA 000 570

College Entrance Examination Board, 1967.

Criticism of Standardized Tests and Testing[1]

David A. Goslin

Standardized tests have been a source of considerable controversy in recent years. Growing competition for jobs, for admission to college, and for educational opportunities in general has led to an intensified search for better ways to evaluate individual abilities and aptitudes and to identify intellectual potential at earlier ages. As a result, standardized tests of various types are being used increasingly at all levels of the educational system and in many other areas of society including the military, the civil service, and business and industry.

This greater reliance on standardized tests has led a number of scholars and others to raise questions about the validity of the tests being used and about their effects on those who take them and on the society that uses them to differentiate among its members. Thus far there have been very few, if any, attempts to bring together all the criticisms that have been leveled against tests and to place them in an analytical framework that would permit a systematic evaluation of their validity. It is the purpose of this paper to provide such a framework and to summarize the major criticisms of tests within this organizational scheme.

I. A framework for discussion

Criticisms of testing must be viewed in the context of three variables: type of test, how it is used, and assump-

---

1. A slightly condensed version of this paper appeared in Science, Vol. 159, February 23, 1968, pages 851-855.

tions regarding the validity of the test (whether it measures what it is supposed to measure). Only if each of these variables is kept clearly in mind from the outset can the force of specific criticisms be evaluated. Unfortunately, as with any such set of variables, it is not always possible to define the boundaries of the argument in an absolute way. But the attempt must be made. A description of the three variables follows.

The type of test being used must be considered first. At the outset a distinction may be made between what I have called ability tests on the one hand, and personality and interest tests on the other.[2] Ability tests may be further divided into two parts: (1) intelligence or aptitude tests, which attempt to measure inherent capabilities or potential of individuals (or, as testers now like to put it, abilities that are acquired over a long period of time), and (2) tests that are designed to measure specific achievements.

Whether one conceives of intelligence and aptitude tests as measuring inherited potential or merely as measuring the general intellectual skills an individual has acquired over the course of his life, the implicit assumption exists that such tests measure a relatively deep and enduring quality. This quality may be viewed as changeable, but it is the expectation that startling changes are rare except under certain specific conditions, such as extreme cultural deprivation. Because of this, intelligence and aptitude tests generate a relatively high degree of anxiety on the part of those who are tested. The high cultural value placed on intellectual abilities in our society also makes any instrument which purports to measure general intellectual abilities a source of fascination. For these reasons, such tests have

---

2. David A. Goslin, The Search for Ability. New York: Russell Sage Foundation, 1963, pp. 13-15.

been a major source of controversy and debate.

Achievement tests -- although less likely to be perceived as unfair, since they are explicitly designed to measure skills acquired over a short period of time in a particular area -- potentially exert a considerable influence on what is taught and how it is taught, as well as on the kinds of skills invividuals choose to acquire. Among all tests, they are distinctive in that it is much easier in the case of an achievement test to see what one is measuring, since the universe of abilities being sampled by the test is theoretically finite and far more easily specified.

Personality and interest tests pose a major problem in that they all depend to some extent upon the honesty of the respondent. Further, since the characteristics being measured may be variously perceived by those evaluating the results of the test as being "good" or "bad" (or "neutral"), there are no clear standards against which to judge performance. Lacking such standards, the person taking the test is in the position of having to decide what the person giving the test is looking for and, indeed, whether it even makes a difference, since to some extent it is possible for the respondent, especially if he knows something about such tests, to create a "false" impression of himself. Therefore, many such tests turn out to be tests of role-playing skill. (They are ability tests, after all!) Section V of this paper will briefly discuss personality and interest tests, but the bulk of it will be devoted to criticisms of ability tests.

The second variable to be considered is the use to which the test is put. All tests may be used in one of two ways: selection and allocation, or counseling -- and sometimes both. Criticisms of tests must be viewed in the light of each of these categories of use. While the distinction is clear in the abstract, confusion is introduced by the fact

that a given test is often used both for selection or allocation and for counseling. Nevertheless, criticisms should be viewed as "use-specific" in most cases.

Whenever a test is used to select among a group of candidates for a position or among applicants for admission to a school or college, or when a test is used to allocate individuals among groups having different characteristics (for example, tracks in a school), the test is serving an essentially predictive function. A prediction is being made about the subsequent performance of the individual relative to the subsequent performance of those against whom he is being compared, and this prediction about his subsequent performance plays a part in decisions that are made about him by others.

Tests may also be used as a basis for providing an individual with information about his abilities, aptitudes, and the like. This use of tests is theoretically different from the previously mentioned use in that the information that is provided the counselee is intended to make it possible for him to arrive at a decision as to his future course of action. In the former case, although a decision is sometimes necessary on the part of the individual (for example, whether to apply), the ultimate decision is made for him by others. It should be noted, however, that counseling may, and frequently does, take the form of directing the individual into one of a number of alternative paths. In this case, depending on the kind of information transmitted and the way it is transmitted, the counselor may actually function as the decision maker. Here again, although we can point to a conceptual distinction, our distinction in all cases is not a perfectly clear one.

Finally, criticisms of tests can be divided into those that raise questions about the validity of tests, and those that have little or nothing to do with whether the

test measures what it is supposed to measure. The criterion for allocating criticisms to one or the other of these categories is as follows: Is the force of the criticism affected by whether we assume the test to be a valid measure of what it is supposed to measure, or not?

II. The extent of testing and test use

No attempt will be made here to describe in detail the extent of testing in the United States, or to provide an analysis of all the ways in which tests are currently used. A number of surveys of test use have been made in recent years, and together these studies provide an adequate summary of the situation.[3] However, the following points are in order.

It is apparent that a great many standarized tests are being given, and there is some evidence that the number is still increasing, although there appears to have been a reduction in the rate of increase during the last several years. Among other things, the provision of funds for the purchase of standardized tests under the National Defense Education Act served as an important stimulus to local school systems, as well as states, to initiate testing programs. Most school systems now make provision for testing pupils at regular intervals beginning in the third grade. Continued increases in the volume of tests given can be anticipated as a consequence of growing school enrollments. But, except for possible new developments such as the National Assessment Program, it would appear that the major growth is over.

---

3. See, for example, David A. Goslin, Ibid., Chapters III, IV, and V; and David A. Goslin et al., Testing in Elementary Schools. New York: Russell Sage Foundation, 1965.

Testing for college admissions has received the greatest attention from the public, as well as from school officials. It should be made clear, however, that College Board-sponsored tests, the American College Testing Program, and the various scholarship testing programs constitute only a very small fraction of the total volume of standardized tests given in schools. In addition, substantial use is made of standardized tests in the military services and in business and industry. Considering only school testing, a very strong case can be made for the fact that standardized tests given in elementary schools have a potentially greater impact on both pupils and schools than do college admissions tests.

Personality tests have not yet been used extensively in schools or colleges, except on an individual basis. The various surveys of test use indicate that only rarely do elementary or secondary schools administer personality tests to groups of students. Where this does occur, however, it deserves special attention, since these tests have been widely criticized, both by the public and by psychometricians. There is evidence, however, that personality tests <u>are</u> widely used in evaluating candidates for positions in business and industry. The Russell Sage Foundation is currently supporting two studies that, it is hoped, will provide information on the general question of personnel selection in business and industry and on the role tests play in this process.[4] At the present time, however, very little of a systematic nature is known about testing at this level.

---

4. These include a study of personnel selection in business and industry being conducted by Stanley Udy Jr. and Vernon Buck of Yale University, and a study of the test-publishing industry under the direction of Milton G. Holmen and Richard F. Docter at the University of Southern California.

Interest tests, on the other hand, have been used extensively in schools, mainly in connection with efforts on the part of schools to provide vocational guidance for pupils. Given the fairly well-documented lack of correlation between scores on interest tests and actual performance in later life, this practice deserves attention.

Although it is possible to count how many tests are given in a school or school system, it is much more difficult to determine how much reliance is placed on test scores, either in decisions that are made about pupils (selection or allocation) or in determining what advice to give to pupils (counseling). Only in cases where a fixed cutoff point in the distribution of test scores is applied as a criterion for eligibility, admission, and the like can one determine exactly the influence of the tests. Although the testing industry officially abhors the use of cutoff points, they <u>are</u> used in some situations: mostly those in which large numbers of candidates are being evaluated (for example, in the National Merit Scholarship Program) or where a system must resort to highly formalized procedures in order to avoid pressures for special consideration.

The difficulty of determining the influence of test scores in the evaluation process is exemplified by the case of the college admissions officer, who typically takes many factors into account in deciding who shall be admitted, and who operates with only a rough formula for the weight to be given to each of these factors. Similarly, the administrator who must decide which children are to be allocated to the fast track and which to the slow track makes use not only of test scores but of a variety of other information, including subjective evaluations of the pupils' potential and so forth. In all these cases, what actually happens may be at considerable variance from what the admissions officer, the administrator, or the guidance coun-

selor says happens. Thus data on the use of test scores, as opposed to the extent of test-giving, are extremely difficult to come by.

It should be noted here that it seems likely that a great deal of the anxiety about testing may be due to the lack of definition of the part tests actually play in the selection or allocation process.

The situation is further confused, as I have noted, by the fact that standardized tests frequently have multiple functions in a school system. Even a test like the College Board's Scholastic Aptitude Test may serve an important guidance function (in helping a student decide where to apply), although its major use is in connection with the admissions process. Among the uses of standarized tests that are commonly reported by high school principals are: (1) as a basis for homogeneous grouping or tracking, (2) as a basis for determining a pupil's strengths and weaknesses so as to make remedial work possible, (3) as a basis for providing the pupil with information about his abilities, (4) as a basis for evaluating the effectiveness of teachers, and (5) as a basis for evaluating the appropriateness of curriculum materials and overall school effectiveness.[5] In evaluating criticisms of tests, the multiple functions of tests should be kept in mind, and any recommendations concerning modifications of testing practices should be phrased in terms of the uses to which tests should or should not be put.

---

5. See Orville G. Brim Jr. et al., The Use of Standardized Ability Tests in American Secondary Schools; and David A. Goslin et al., Testing in Elementary Schools. Both New York: Russell Sage Foundation, 1964.

III. Criticisms of the validity of ability tests

The following criticisms of testing are subject to assumptions about the validity of ability tests: that is, do ability tests measure what they are supposed to measure, and is what they measure relevant? In the case of each of the criticisms discussed in this section, it may be assumed that the force of the criticism is substantially affected by whether or not tests are determined to be relatively accurate predictors. In some cases, the criticism focuses directly on the problem of validity itself.

A number of critics of standardized tests have claimed that tests, as currently designed, are unfair to certain individuals or groups because of the characteristics of the tests themselves. These critics suggest that, for some groups, test scores are not valid predictors of the subsequent performance of members of the group. Three types of individuals have been singled out in particular by these critics.

The first group for whom it is claimed tests are unfair are the deep thinkers. Critics who take this position claim that certain items on standardized tests penalize a bright student because of the fact that they are ambiguously worded, or because the alternatives presented include one or more options (scored as incorrect) which a mediocre student passes by but which an extremely bright student correctly perceives as being possibly correct answers.[6] One cannot dispute the fact that Banesh Hoffmann and others have demonstrated clearly the existence of such items on tests currently in use, including the College Board's SAT.

Although no major studies have been done to determine

---

6. See, for example, Banesh Hoffmann, The Tyranny of Testing. New York: Crowell-Collier, 1962.

whether any extremely bright students have actually suffered as a result of poorly written tests, Hoffmann's point remains valid, at least in the abstract. It seems unlikely to me, in our achievement-oriented society, that very many geniuses remain undiscovered, regardless of their performance on standardized tests (or, more important, that more geniuses are missed because of standardized tests than would be missed with alternative selection techniques). Nevertheless, this criticism requires empirical examination.

Incidentally, this criticism underscores an inherent difficulty with high-level objective tests: in order to make such a test difficult enough to differentiate among a group of able candidates, it is virtually impossible to avoid some ambiguity in item construction. This point has relevance for what I think is a more serious criticism of tests (which is discussed below): their general imperfection as predictors.

A second group for whom it is claimed tests are unfair are the culturally disadvantaged and members of distinctive cultural groups. Almost by definition, any test that is designed to be given to a broad spectrum of individuals in our heterogeneous society must discriminate against those individuals whose cultural background is different from that of the majority. To take an extreme case, if a pupil cannot read English because Spanish is spoken at home, he is not likely to do well on an English reading comprehension test or, more important, on any test in which the ability to read English is either a part of the test or is an assumed prerequisite for understanding what one is supposed to do on the test. This principle holds not only for such extreme cases but for the members of any group whose life experiences differ significantly from those on which the test was standardized.

This problem is partly one of standardization. Conceivably, special norms could be developed on any test for

every distinctive group that is likely to take the test, in such a way that both inter- and intragroup comparisons could be made. But another principle is involved.

Standardized tests are designed in most cases to predict success of individuals in the broader society or in the setting to which the individual wishes to gain admission. From this standpoint, it can be argued that tests are doing the job expected of them if they do discriminate among members of different groups. If we assume, for example, that the ability to speak English with facility is necessary for success in our society, then a test of verbal ability based on facility with English is not an unfair yardstick to apply to individuals who do not possess this ability, whether they come from a foreign-language background or are members of a culturally deprived group. In such cases, it can be pointed out that it is not the test that is unfair but rather the circumstances that have permitted the deprivation to persist.

The latter argument is reasonable enough; however, it is not entirely satisfying, especially to members of disadvantaged groups. Part of the difficulty lies in the assumptions that are made about an individual as a consequence of his poor performance on a test due to an inferior or different background. If his poor performance is interpreted as indicating a general lack of intellectual ability, as opposed to specific deficiencies in the skills demanded by the tests, it is not the situation but rather the test and its interpreters who are being unfair. But how do we decide whether his poor performance is due to a general lack of ability or to cultural deprivation? There is no satisfactory answer to this question at present. Intragroup norms may help. Correction factors might be introduced on a formal basis for members of certain groups (no doubt this is now being done informally). All in all, it seems likely

that any inferences whatsoever about the general intellectual abilities of members of disadvantaged or other special groups on the basis of test scores should be avoided -- especially when those being tested are young.

Finally, some critics have pointed out that tests may be unfair to individuals who lack experience in taking standardized tests. Taking a standardized test requires, in and of itself, special skills. It may be assumed that for almost everyone these skills are developed as a result of repeated contact with objective types of tests. Some individuals, however, have an opportunity for greater contact with tests than others. It is unknown just how much experience with standardized tests is necessary before this factor becomes of negligible importance in influencing the individual's test performance in relation to the performance of others. It may be assumed, however, that tests are "unfair" to some individuals who have not had the requisite experience in dealing with tests of this sort. From this point of view, the extensive testing being done in elementary and junior high school is beneficial, but we know that the amount of experience various individuals have had with tests still varies considerably from locality to locality and from school to school. The problem is particularly acute for individuals who come to this country from places where such tests are not widely used (for example, foreign applicants to American graduate and professional schools).

The imperfect-prediction problem

Standardized ability tests are not perfect predictors of subsequent performance, even in situations that require abilities which are very similar to those required on the test. Highest coefficients of correlation between test scores and measures of subsequent performance are obtained

for short-range academic performance: for example, twelfth-grade standardized test scores predict first-year college grades fairly well.[7] But, as the length of time between the test and criterion situation increases, the magnitude of the correlation is reduced. Similarly, as the criterion situation becomes more dissimilar to the test situation, the correlation is reduced. Thus, most existing studies show no correlation between test scores and subsequent occupational success (nor is any correlation shown between academic performance as measured by grades and subsequent occupational success). Given the fact that test scores correlate only moderately with long-range academic performance and not at all with postacademic performance, serious questions are raised about the usefulness of such scores and the amount of reliance that ought to be placed on them. Three factors that contribute to this lack of correlation require explication.

First, there is the problem of range restriction. It is obvious that accurate predictions about the relative performance of individuals are more easily made where there are sizable differences between individuals: a high degree of variance in the distribution of abilities being measured by a test makes prediction easy. On the other hand, where differences between members of the group being tested are small, predictions of later performance of the members of the group relative to one another are extremely difficult to make. Thus when one is attempting to make predictions within a relatively homogeneous group, such as college graduates, such predictions are of necessity bound to be extremely risky. The phenomenon of range restriction

---

7. For a summary of research on the prediction of academic performance, see David E. Lavin, The Prediction of Academic Performance. New York: Russell Sage Foundation, 1964.

no doubt accounts in large part for the lack of correlation between either test scores or academic performance and occupational success among able students.

Second, there is the problem of assuming the existence of a linear relationship between qualities measured by a standardized test and occupational success. Simplistic notions about the role of intellectual abilities in individual achievement suggest a linear relationship between such abilities and success: the more intelligent one is, the more likely he is to succeed. However, a variety of studies have indicated that the relationship between intellectual abilities and success in our society is far more complicated than is suggested by such a model.

For example, although Lewis M. Terman demonstrated clearly that his gifted group as a whole was more successful than less intellectually able groups, he found no relationship between intelligence and later performance within the gifted group.[8] These findings are corroborated by the previously noted lack of correlation between college performance and subsequent nonacademic success. All this suggests that, at least in the context of the way our society currently operates, intellectual abilities may function as a threshold variable in relation to occupational advancement. It may reasonably be hypothesized that a minimum level of intelligence is required for most occupations, but once at or above this minimum, an individual's achievement relative to others in the same field will be determined by other qualities not measured by tests of intellectual abilities.

It should be noted that there are no doubt differences between fields of endeavor, not only with respect to the minimum level of intelligence necessary but also with re-

---

8. The Gifted Group at Mid-Life. Stanford, Calif.: Stanford University Press, 1959.

spect to the amount of difference made by increments over this level in one's chances of achieveing success. Another way of stating this is to say that qualities other than intelligence are more important in some fields than in others. (Incidentally, this does not have to be the case; it just happens that our society works this way at present. One could, for example, imagine a society in which a perfect correlation between intelligence and success could be achieved by assigning all jobs and status in the society on the sole basis of intelligence.)[9]

Finally, the qualities not measured by standardized ability tests raise difficulties. Standardized ability tests measure only a few of the large number of personal qualities and characteristics that are important for success in our society. Motivation, creativity, social skills, physical appearance, and a variety of other characteristics contribute in varying degrees to one's chances of achieving success. In addition to explaining a lack of or low correlation between test scores and subsequent performance, this fact underscores the conclusion that excessive reliance on tests in the selection process may result in potentially highly successful individuals being overlooked.

All the foregoing points suggest the need for increased efforts to develop standardized measures of a much wider variety of personal qualities, including abilities, than are currently available. As will be noted below, reliance on a single criterion, such as intellectual ability, in the allocation of status runs the risk of generating a rigid class structure that may have unfortunate consequences both for individuals and for the society as a whole. Of course, the observed low correlations between

---

9. Just such a society is envisioned by Michael Young in his book The Rise of the Meritocracy. London: Thames and Hudson, 1958.

test scores and subsequent occupational performance indicate that our society does make use of many characteristics other than intellectual ability in allocating status at the present time. However, the only formalized measures we have are of intellectual skills. As a result, these measures are often viewed as being more important. In our efforts to eliminate subjectively of judgment and thereby promote "equality of opportunity," we are forced, paradoxically, to rely more heavily on measures of intellectual skills -- measures that, in turn, may lead to a higher degree of stratification.

The rigid use of test scores

A major source of criticism of standardized ability tests is the practice of rigid reliance on test scores in evaluating candidates for positions, college admission, homogeneous grouping, and the like. A typical "horror story" advanced by the critics takes the following form:

Johnny gets an IQ score of 78 on a third-grade group intelligence test and is therefore placed in the section for retarded children. His parents appeal this decision on the grounds that Johnny has always done well in school up to that point, and that they can detect no signs of retardation at home. Johnny may be given an individual IQ test by the school psychologist, the result of which "confirms" the group-test score. The parents' appeal is rejected by school officials, who point out that school regulations require that any child whose IQ is below 80 must be placed in the section for retarded children. Subsequently it is discovered that Johnny is terrified of school or school psychologists, or has a special reading problem, which resulted in his abnormally low score, and that in fact he is above average in general intelligence. But discovering this takes two years, during which time Johnny is treated as a retarded child.

When faced with such stories, professionals typically respond by pointing out that (1) the school psychologist was at fault for not identifying Johnny's special problem, or (2) the system is at fault for using test scores in such a rigid way, or (3) it is unfortunate, but mistakes are sometimes made. However, none of these excuses is very satisfying to Johnny or his parents.

Almost everyone can cite similar cases at all levels in the system. In his book, They Shall Not Pass, Hillel Black devotes an entire chapter to a case similar to the one described above. (He calls it: "How They Buried Maria.")[10]

Several general points may be made on this issue. First, while testers acknowledge that occasional errors result from the use of standardized tests, they claim that fewer errors are made with tests than with alternate techniques. Second, testers point out again that it is not so much the tests that are at fault as the system that uses them in inappropriate or overly formalized ways. They point to their repeated attempts to get test users to recognize the fact that a test score is not an absolute indicator of ability and that test users should take into account other factors in addition to test scores. Finally, however, because tests generate a numerical score there is an inherent tendency to perceive that score as a precise measure and to make distinctions between individuals on the basis of small differences in their respective scores in the absence of other quantifiable indicators. Formalistic reliance on test scores, for example, is an easy way to protect oneself against demands for special favors.

Inherited versus acquired abilities

Another basic criticism of standardized ability tests and, more important, of the ways they are used involves the

---

10. New York: William Morrow and Company, 1963.

question of what assumptions those who use tests make about the qualities the test is measuring. This boils down to whether it is assumed that the test measures innate capabilities (which are therefore presumed to be unchangeable), or whether it measures learning. Few people who possess any sophistication in psychometrics these days take the position that even intelligence tests measure only innate capabilities.

However, our data indicate the existence of significant differences in opinion on the question of whether the qualities measured by intelligence tests are more or less influenced by learning than by inherent potential. Whether one views a test score as providing an accurate measure of an individual's innate ability is likely to have an important effect on the use one makes of test scores. If one views a child's performance on a test as being influenced primarily by what he has learned, as opposed to his innate capability, then one is less likely to make long-run predictions about the child's ultimate success on the basis of his test scores (after all, his motivation might increase, and he might do better next time). Since much hostility to tests stems from the fact that many people resent having their children or themselves classified as potential successes or failures, clarification of this issue is badly needed.

The self-fulfilling prophecy problem

One of the most important criticisms of tests is that they contribute to their own validity of functioning as self-fulfilling prophecies. It is hypothesized that a child who does well on a test and, as a consequence of his performance, is placed in an advanced class or receives special attention from his teachers, or who is admitted to a good university, is more likely to do well in the future than

the child who initially got a lower score on the test. The likelihood that the optimistic prediction made on the basis of a high test score will be fulfilled is therefore increased by the fact that special advantages are given to the person who got the high score on the test. The same condition may be made in the case of an individual who does poorly on a test and as a consequence is denied opportunities.

Experimental data based on a study conducted recently by Robert Rosenthal[11] tend to confirm the validity of this hypothesis. Rosenthal gave all the children in four California elementary schools an ordinary intelligence test at the beginning of the school year. He informed the teachers that the test he had given was specially designed to identify children who could be expected to show substantial IQ gains during the coming year. In each class, he then selected at random 10 children and informed the teachers that these children had done particularly well on the test. This group in each class formed the experimental group, and the remainder of the children in each class served as the control group.

An intelligence test administered at the end of the school year showed that the experimental groups in grades kindergarten, one, two, and three had made significant gains in IQ when compared to the children in the control groups. In addition, teachers rated children in the exper-

---

11. Robert Rosenthal and Lenore Jacobson, "Self-Fulfilling Prophecies in the Classroom: Teachers' Expectations as Unintended Determinants of Pupils' Intellectual Competence." In M. Deutsch et al., Social Class, Race, and Psychological Development. New York: Holt, Rinehart & Winston, in press. See also Rosenthal and Jacobson, Pygmalion in the Classroom: Teacher Expectations and Pupil Performance. New York: Holt, Rinehart & Winston, in press.

imental groups as being better than those in the control groups on a variety of personal qualities such as cooperativeness, interest in school affairs, social ajustment, and the like. The findings indicate clearly that teachers' expectations contributed substantially to the intellectual growth of the children in the experimental groups. In this case, the initial test score reported to the teachers turned out to be a self-fulfilling prophecy. The implications of this point are far reaching, especially for policies on the use of standardized intelligence tests in the elementary grades.

IV. Criticisms that are independent of the validity of tests

The following criticisms may be hypothesized to hold whether one argues that tests are valid measures of ability or not. In some cases, the force of the criticism is increased if one assumes tests to be highly valid predictors. These criticisms, therefore, stem from the potential social effects of testing, as opposed to questions regarding the accuracy of tests.

We have noted that standardized ability tests are used throughout the educational system and that children take such tests at periodic intervals. In addition, the spread of the technology of standardized test construction has led many teachers to make use of objective items in tests they themselves construct for their students. It has been suggested that continual exposure to multiple-choice items during the elementary and secondary grades tends to result in an unfortunate constriction in the ability of children to reason. In particular, it is claimed that emphasis on evaluation techniques in which there is always a right and wrong answer makes it difficult for children to acquire

the ability to deal with issues on which there is no clear right or wrong answer. Children, it is claimed, are therefore handicapped when they attempt to work through questions involving ethical or philosophical judgments, or when arriving at a decision depends on identifying the assumptions one is going to begin with.

To the best of my knowledge, no data exist to tell us whether or not this is a valid criticism of tests. Colleges claim that incoming students do not write as well as they used to, but there is no way of knowing whether this complaint is just a case of the older generation complaining about the new one or, if true, whether it is owing solely to the fact that a larger proportion of students goes to college today. Research on this topic is difficult, since it is getting harder to find a control group (that is, comparable students who have not been exposed to standardized tests). Nevertheless, the argument has a certain logic and requires careful consideration.

When a student takes a college entrance examination -- or almost any standardized test, for that matter -- not only he but his teachers and his school are also being tested, since his performance reflects in part the adequacy of the training he has received. As a consequence, it may be argued that tests have a potentially significant impact on what is taught in schools and the way it is taught. Our data indicate that only a very small minority of teachers say that they spend any significant proportion of time preparing students to take standardized tests, or that they have ever altered a course they teach because they found out that the subject matter covered by a standardized test was different from what they normally taught. Nevertheless, there is evidence that in many situations standardized tests do exert an influence on what is taught. The famous case of the New York Regents examination pro-

gram is pertinent here. Since both teachers and schools were being evaluated along with students, there was, and still is, considerable pressure to prepare students to take the Regents achievement examinations. Reports of students' being drilled on old copies of the Regents examinations were common. That tests have had an impact on the curriculums in this case cannot be disputed.

Whether or not teachers make special efforts to prepare students for taking particular standardized tests, such tests can have a more general impact on curriculums. For example, widely used external examinations, like the College Board's Achievement Tests, may result in pressure on a school system to adopt a new curriculum if the school perceives that the content covered by the test differs significantly from that being presented in the school. Thus, standardized tests based on the new mathematics curriculum can be expected to speed the adoption of this curriculum in schools.

It should be noted that such an effect is not necessarily deleterious. Standardized tests may serve to raise school standards as often as they function to set limits on innovation and experimentation. This was the idea behind the Regents examination program when it was initiated. The problem, of course, is striking a balance between raising standards and setting arbitrary limits.

Increasingly, schools, colleges, and testing agencies are following the practice of providing individuals with information about how well they did on standardized tests. This information may be given in the form of a specific score or percentile rank, or it may be presented in more general terms. Regardless of the way in which such information is transmitted to the examinee, it may be hypothesized that it will have some effect on his self-image and, in turn, an effect on his motivation, aspirations, and so

forth. The effect of receiving information about one's abilities will depend on a variety of factors, including the legitimacy of the source of the information, the perceived accuracy of the test, and the degree to which the information confirms one's own estimate, including the extent to which it is threatening or rewarding. Obviously, individuals make use of many different types of information in arriving at an estimate of their abilities: standarized test scores are only one of many ways in which individuals get information about their capabilities. Data from a national sample of high school students indicate further that test scores are of relatively minor importance in shaping self-estimates of ability in comparison with such things as school grades, comments made by peers and parents, and relationships with teachers.[12]

Nevertheless, test scores have a potentially great impact in particular instances where an individual's self-estimate is at considerable variance with the record of his performance on the test and where rationalizations of poor performance are unavailable, or where the score is substantially higher than his own estimate. Under such conditions we may expect a shift in self-estimate of ability to affect the individual's aspiration level, his motivation to achieve, and, secondarily, personal decisions with respect to future courses of action.

At a broader level, it might be helpful to consider the consequences for overall aspiration levels in the society of a system in which individuals were classified very early with respect to their abilities and available opportunities for the future. Thus far, very few data have been

---

12. Orville G. Brim Jr. et al., The Use of Standardized Ability Tests in American Secondary Schools. New York: Russell Sage Foundation, 1964.

collected on this issue, and it is certainly one that would profit from further research. In addition, rather complicated ethical and moral considerations are raised, whatever the research findings might be.

The use of any single criterion or set of criteria to sort individuals into groups or to decide which individuals will be admitted to a group has important implications for the structure and characteristics of the groups thus formed. These implications may be examined under the following headings: (1) social structure within groups, (2) tendencies toward uniformity in the characteristics of group members, and (3) implications for the society as a whole.

1. Implications for the social structure of groups: Standardized tests are currently being widely used as an important criterion for allocating students to instructional groups or tracks within schools. The result of this use of tests is social differentiation within schools based to some extent on the qualities measured by standardized tests. To the extent that schools organize pupils according to their abilities, possibilities of social contact between children of differing levels of ability (as measured by standardized tests) are reduced. Research indicates that such differentiation within schools may have a negative effect on the performance levels of low-ability pupils while not significantly facilitating the performance of high-ability pupils.[13] In addition, it is clear that ability grouping impedes the process of acculturation of members of culturally deprived groups, who tend to end up together in the low-ability groups. Finally, since the school plays a major role in courtship and mate selection

---

13. Alfred Yates, Grouping in Education. New York: John Wiley & Sons, Inc., 1966.

by providing an important setting for contacts between boys and girls during adolescence, subgroup differentiation within the school can be expected to affect to at least some degree the process of mate selection.

2. Implications for groups that use tests to select their members: To the extent that any group relies heavily on a single criterion for selecting its members, whether that criterion is ability or something else, there will be a strong tendency toward uniformity in the characteristics of members of the group. One of the problems faced by our elite colleges and universities, for example, is how to achieve diversity in the student body while admitting only students of exceptional ability. The problem becomes more acute when standardized tests are heavily relied on as a measure of intellectual ability. This point is closely related to the following problem.

3. Implications for society as a whole: Dael L. Wolfle has pointed out that the success of modern, complex societies depends in large part on the availability of a talent pool in which a great diversity of abilities and skills is represented.[14] In order to create such a talent pool, rewards in the form of social status, prestige, and economic returns must be provided for individuals possessing many different talents. A tendency to rely heavily on standardized tests of a more or less limited set of intellectual skills in the allocation of opportunities for the achievement of social status must necessarily result in a reduction in the diversity of talent available. This is partly a problem of devising tests designed to measure a greater range of abilities than are measured by current tests. It is also a problem of ensuring adequate rewards for indi-

---

14. "Diversity of Talent," American Psychologist, Vol. 15, August 1960, p. 539.

viduals who possess abilities that are not measured by tests but that are important for the successful functioning of the society.

A test is a potential invasion of privacy in the sense that it makes information about a person available to other persons. Very important values in American society suggest that it is a basic right of individuals to decide to whom and under what conditions they will make available information about themselves. Correlative to this point, however, is the fact that participation in the society carries with it certain obligations and responsibilities. Further, the right of groups to demand information from those who aspire to enjoy the privileges of group membership is clearly understood. Thus, no one is likely to object to being given a driving test before being permitted to operate a motor vehicle. Similarly, few people object to the requirement that they must take an entrance test in order to gain admission to a university or college. In each of these cases, the right of a group to information that is relevant to its stated objectives and goals about candidates for membership has been established beyond question. Two important questions remain, however.

First, under what conditions does a group have the right to ask aspiring members for information that is irrelevant to its purposes and goals (and how does one decide what is relevant and what is irrelevant)? In order to answer this question, it is probably necessary to make a distinction between public and private groups. It seems reasonable to assert that a private group has the right to ask applicants for membership anything it wants to ask them, relevant or irrelevant. In this case, it is up to the applicant to decide whether he wishes to reveal this information. In the case of a group supported by the society as a whole, including all of the potential applicants to

the group, this is a more difficult question.

Would it be, for example, legitimate for the state to ask individuals to reveal information about their sex lives as a requirement for obtaining a driver's license? Most of us would, I think, object to such a requirement on the grounds that it represents an invasion of our privacy that is not justified by the service being rendered. Just such objections are being raised to the use of personality tests in schools, and there would appear to be a justification for such objections (see below). The issue is one of relevance: must the school have such information in order to do its job?

There is, however, a second and more difficult problem in the case of school testing. In each of the cases presented above, the individual retains a choice as to whether he will submit himself to the test or not. Thus, if an individual does not want to take the College Board's Scholastic Aptitude Test, he does not have to. Nor does he have to submit to a driver's test. As a result of his decision he may have to give up his chances of attending Harvard or driving an automobile, but the choice in each case is his. But for the most part parents do not have a choice about whether their child will take tests or not, including standardized tests. A parent might move to a community in which the school system did not use standardized tests (if he could find one), or he might send his children to a private school that did not administer tests (if he could afford one). But for most parents these are not realistic alternatives. Thus, most children or their parents have no choice about whether they will allow others to determine how intelligent they or their children are through the use of standardized tests. Children must go to school, and they must take tests in school.

Does this constitute an invasion of privacy? Carried to

its extreme, an affirmative answer leads one to the conclusion that children should be permitted to refuse to take tests given by their teachers in class. Although this sounds absurd, it is not an unreasonable claim. If a child refused to participate in classroom tests, he would fail his courses and probably not be promoted, but this would be his (or his parents') decision. Few would argue that the school does not have a right to require pupils to demonstrate their proficiency in school subjects before according them advanced status. The more difficult question is whether it is also the right of the school to require pupils to demonstrate their general intellectual ability apart from their proficiency in reading or mathematics or social studies. If a child refused to take an IQ test given in school, would he fail his courses? Does a school need such information in order to decide whether a child should be promoted?

One additional point may be raised here. Assuming one concludes that a school has the right to collect information from pupils about their intellectual abilities, does the school also have the right to withhold this information from the pupil and his parents? Conversely, what rights do parents and pupils have to know what information the school possesses about them? In at least one case (in New York State) the courts ruled in 1961 that parents do have the right to access to information on the pupil's permanent record card maintained by the school.[15]

Obviously these are extraordinarily complicated issues, which cannot be resolved in this short paper. Nevertheless, they are issues that I suggest deserve serious con-

---

15. Van Allen v. McCleary, 27 Misc. 2d 81, NYS 2d 501 (Sup. Ct. Nassau Co. 1961).

sideration in the context of any discussion of the role of standardized testing in our society.

V. Criticisms of personality tests

The bulk of this paper has been concerned with ability tests as opposed to personality tests. Thus far personality tests have not played a very significant role in our schools, either for selection and allocation, or in counseling, although school psychologists occasionally use personality tests in attempting to understand the causes of unusual behavior in particular cases. There are several reasons for this.

First, strong reactions from parents in those cases in which a school has administered personality tests to groups of children have tended to discourage school administrators from engaging in such practices on any large scale. Second, the scarcity of personnel qualified to administer and interpret personality tests and the amount of time involved in using such tests effectively is an important factor. Third, except in cases of extremely aberrant or disruptive behavior, the gathering of information relevant to children's personality characteristics is peripheral to the main task of the school. Finally, the validity of personality tests has been widely questioned, and considerable doubt remains as to whether they are of any use at all, except perhaps in extreme cases.

Sales figures from commercial test publishers indicate, however, that personality tests are widely used in personnel selection by business and industry, although as yet we have little good data on how they are used in specific terms. This phenomenon deserves careful attention, as does the less frequent use of such tests in schools.

Criticisms of personality tests center on two major

points: the validity and reliability of such tests, and the extent to which they are considered to be an unwarranted invasion of peronal privacy. Within the scope of this paper it is not possible to undertake a detailed analysis of the validity and reliability of personality tests. It is enough to say here that an extensive body of literature exists on the relationship between scores on personality tests and other indicators of performance, social adjustment, and independent evaluations of personality characteristics. In some cases, personality measures have proved to be useful predictors of the subsequent behavior of individuals; in many other cases little or no relationship has been found.

Inconsistencies and contradictions in the research findings may be explained by a variety of factors. These include: (1) the extremely wide variety of types of personality tests and the range of attributes they are designed to measure; (2) the lack of a clear conception of the structure and dynamics of personality itself, which make sampling of personality characteristics extremely difficult; (3) the problem of getting good criterion measures -- that is, reliable independent evaluations of behavior that permit inferences about the role of personality variables in that behavior; (4) the problem of examinee honesty on questions that have no clear right or wrong answer; (5) the problem of standardization; and (6) the problem of reliability in the interpretation of responses, especially on projective types of instruments.

Much more serious questions are raised by personality tests than by ability tests in the matter of personal privacy. Since it is more difficult to demonstrate the necessity of obtaining such information in most situations, and the information itself is likely to be considered to be of a more private nature, much of the public hostility toward personality tests has its source in the belief that such

tests are an unwarranted invasion of privacy. All one has to do is ask a group of school children to answer questions dealing with their sexual fantasies and anxieties, or their feelings toward their parents, and the wrath of the community is invoked in full force. Even tests that do not contain items having an easily identifiable content are likely to be viewed with hostility, since people are afraid of what inferences might be drawn from their responses. All the issues above are relevant here, with added impact due to the nature of the tests and the lack of apparent justification for their use. Under these conditions it is not surprising that personality tests have been more widely used in business and industry than in schools.

Summary and conclusions

Recent debates over the validity of standardized tests and their proper use have usually failed to specify the context within which the criticism is being advanced. This paper has been an attempt to provide a conceptual framework within which to evaluate various criticisms of tests and to summarize the major issues that have been raised in regard to testing. At the outset a distinction was made between criticisms directed at the validity of tests and criticisms that are, in the main, not affected by whether one assumes tests to be valid indicators or not. It was noted further that all criticisms of tests must take into consideration the type of test and the use to which the test is put.

Under the general heading of criticisms of the validity of tests the following issues were raised: (1) the question of the extent to which tests may be unfair to certain groups and individuals, including the extremely gifted, the culturally disadvantaged, and those who lack experience taking tests; (2) criticisms based on the fact that tests are

not perfect predictors of subsequent performance; (3) problems caused by the rigid use of test scores; (4) the problem of assuming that tests measure inherent qualities of individuals; and (5) the question of how much tests contribute to their own predictive validity by serving as self-fulfilling prophecies.

Within the category of criticisms that are more or less independent of test validity, the following were discussed: (1) the effect of objective tests on the thinking patterns of those who are tested frequently; (2) the effect of tests on school curriculums; (3) the effects of tests on the examinee's motivation and aspirations; (4) the effects of tests on groups that use tests as a criterion for selection or allocation or both; and (5) the problem of privacy. The final section of the paper was devoted to criticisms of personality tests as opposed to ability tests.

Several concluding remarks are in order.

This paper has focused almost entirely on <u>criticisms</u> of tests. The positive value of standardized tests should not be ignored, however. In regard to this point it is important to keep in mind the question of what alternative measures might be used if we were to abandon standardized tests altogether.

A major conclusion to be drawn from the preceding analysis is that we must begin thinking about tests in much broader perspective -- one that includes consideration of the social effects of tests as well as their validity and reliability.

Finally, it would appear from the foregoing that an effort should be made to develop rational and systematic policies regarding such things as the use of tests with culturally disadvantaged students, the dissemination of test results, and the problem of invasion of privacy. Such policies can be formulated only if we are willing to take

a long hard look at the role we want testing to play in the society. Standardized tests currently are a cornerstone in the edifice of stratification in American society. It is up to the social scientist to engage in the kinds of research that will enable policy makers in education, business and industry, and government to determine in a consistent and rational way the ultimate shape of this edifice.